

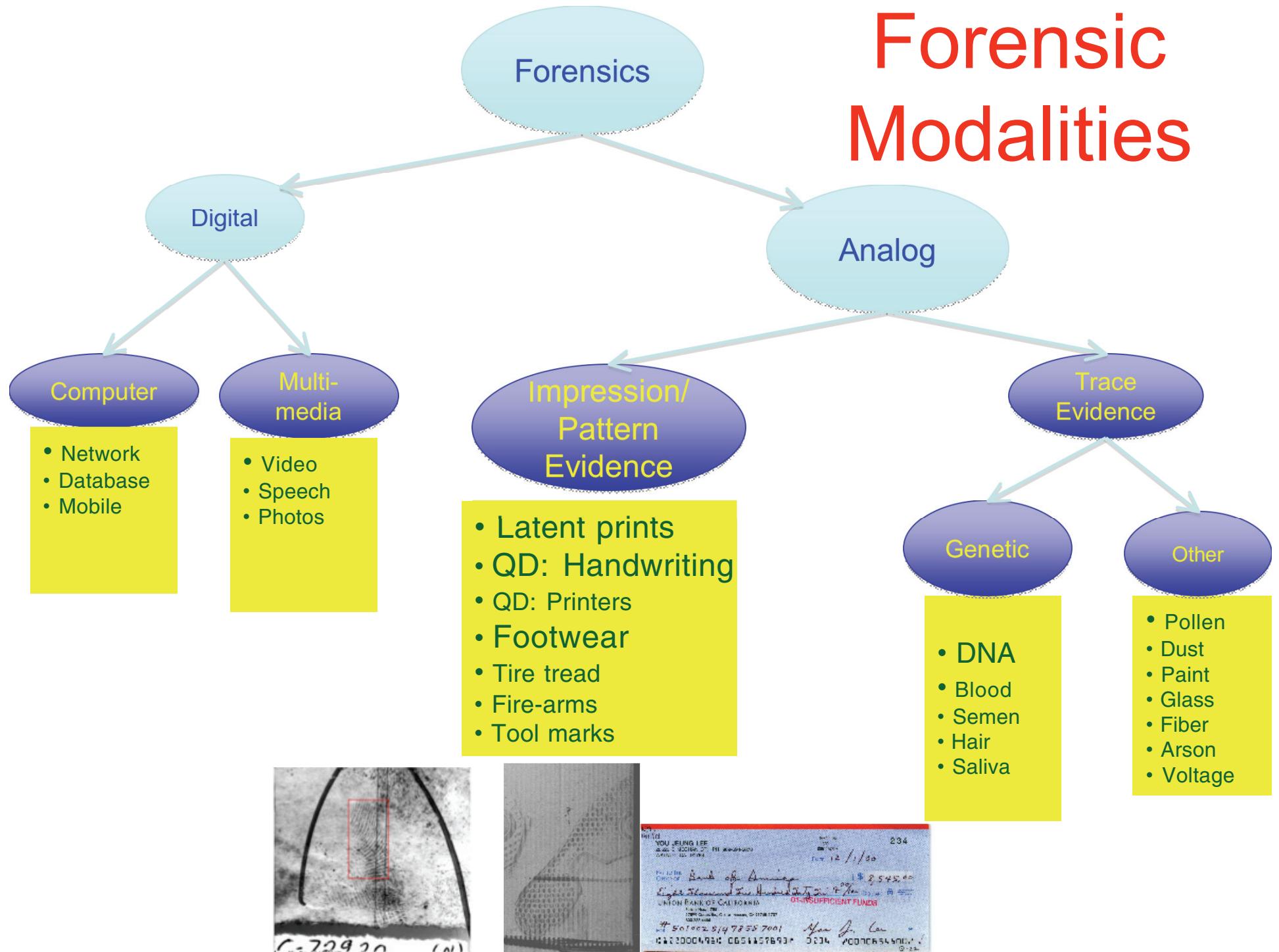
Probability Models for Impression and Pattern Evidence

Sargur Srihari

Department of Computer Science and Engineering
University at Buffalo, The State University of New York

Indigent Symposium 2010
MayFlower Hotel, Washington DC

Forensic Modalities



Forensic Opinion

- Courts allow Latent Print Examiner to opine on ultimate question of individualization
 - Evidence is attributed to a single individual and no other
- Three possible opinions
 - The evidence
 - Individualizes
 - No other individual on earth
 - Inconclusive
 - Excludes
 - Definitely not this individual

Madrid Bombing Case

Latent Fingerprint
on bag of detonator



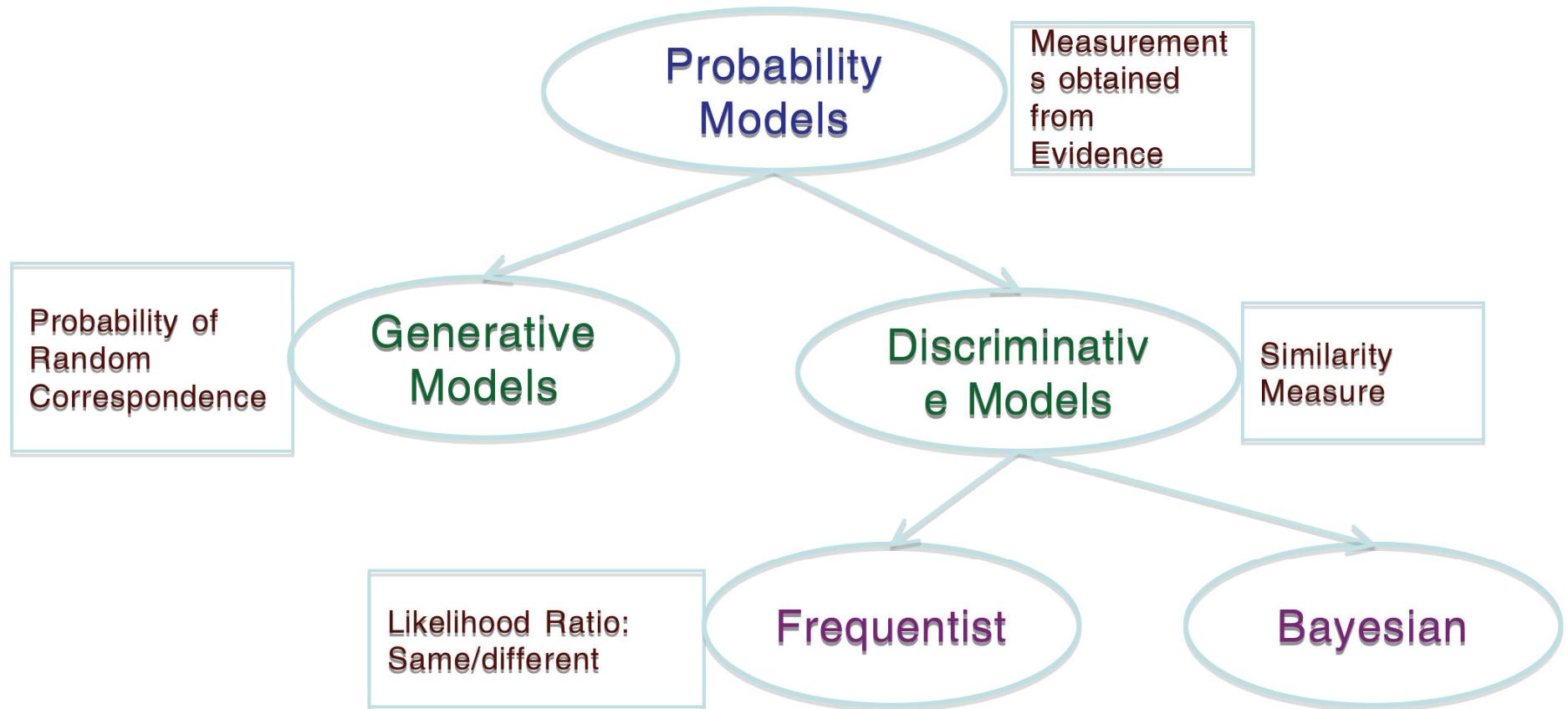
Brandon Mayfield
Identified by FBI



Algerian National
Identified by SNP



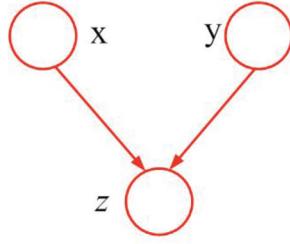
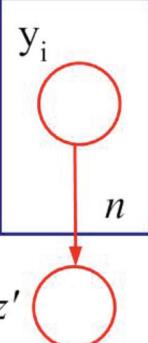
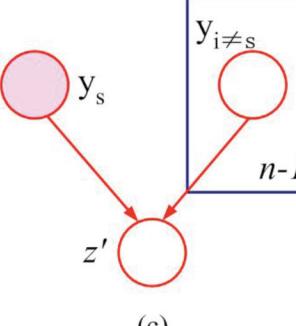
Methods for Expressing Uncertainty



How to Compute Uncertainty?

- Discrete Variables
 - Birthdays
 - DNA
- Continuous Variables
 - Heights
 - Pattern/Impression evidence (Ongoing Research)
 - 5 or 9-point scale suggested by SWGs-- no guidelines
 - New statistical models being developed

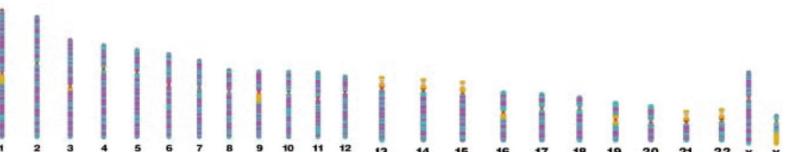
Generative Models: Several Probabilities

	Random two have same birthday ($n=2$)	Some two among n have same birthday	A specific birthday among n
Graphical Model	<p>PRC</p>  <p>(a)</p> $z = \{1, 0\}$	<p>nPRC</p>  <p>(b)</p> $z' = \{1, 0\}$	<p>Specific nPRC</p>  <p>(c)</p> $z' = \{1, 0\}$
Inference	$p(z) = \sum_x \sum_y p(z x, y)p(x)p(y)$ PRC is $\rho = p(z=1)$	$\rho[n] = 1 - (1 - \rho)^{\frac{n(n-1)}{2}}$	$p(z' y_s) = \sum_{\mathbf{Y}'} p(z', \mathbf{Y}' y_s) = \sum_{\mathbf{Y}'} p(z' y_s, \mathbf{Y}')p(\mathbf{Y}')$

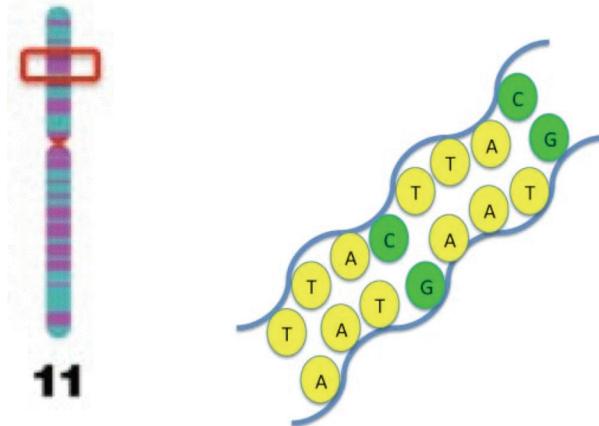
Generative Model: DNA

Genome: sequence of 3×10^9 base-pairs (nucleotides A,C,G,T)

Represents full set of chromosomes



Actual Electron photomicrograph



Single Chromosome: $\sim 10^8$ base-pairs

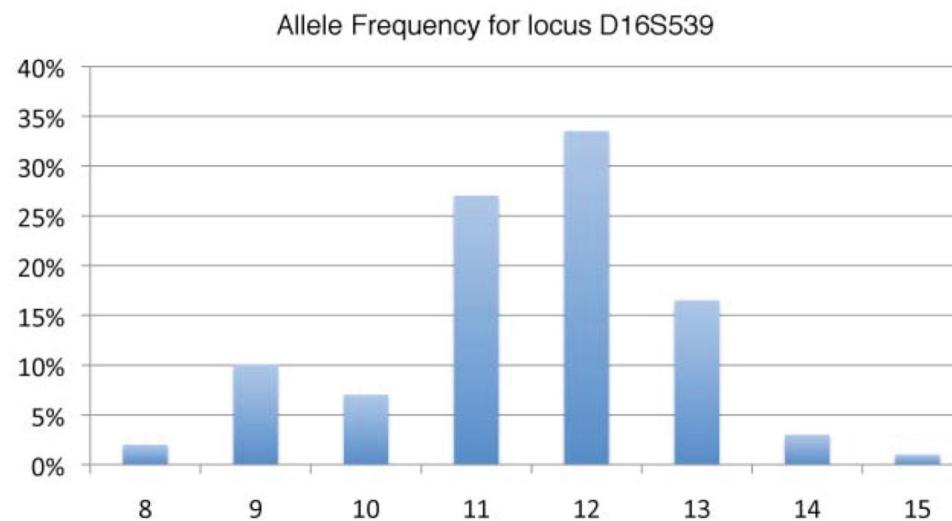
Genome has 46 chromosomes
(22 are repeated plus XX and XY)

Large portions of DNA have no survival function (98.5%) and have variations
useful for identification

TH01 is a location on short arm of chromosome 11:
short tandem repeats (STR) of same base pair AATG
Variant forms (alleles) different for different individuals

Generative Model: DNA

Allele Frequency of single locus for 200 individuals



DNA profile of 13 loci:

Average match probability (PRC) is 0.1 per locus, 10^{-13} for a profile

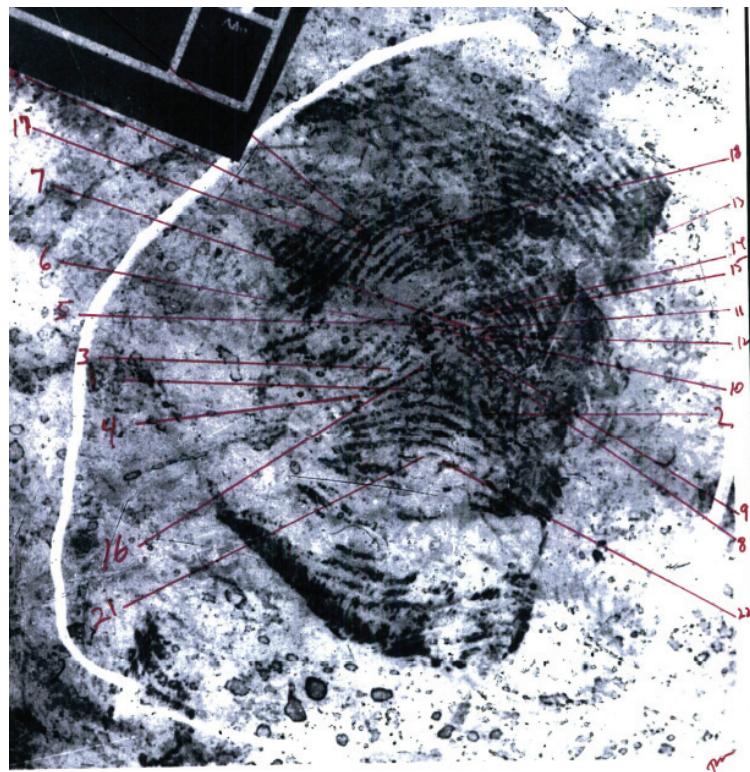
If database has 1 million entries,

since there are 500 billion pair-wise matches, $nPRC = 0.05$

However specific $nPRC$ can be much lower

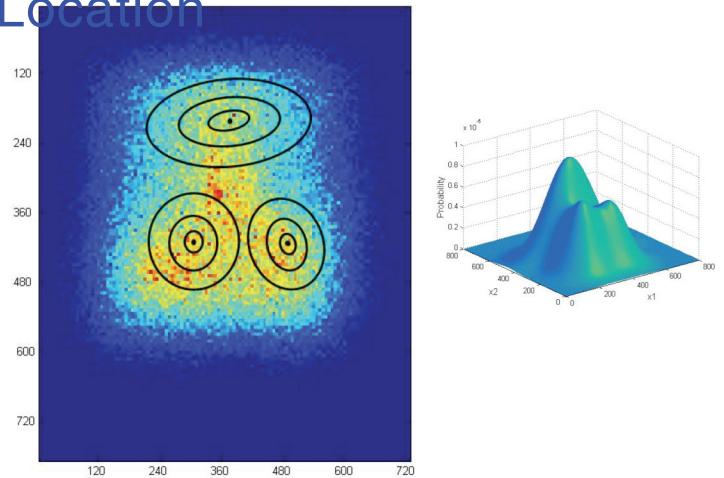
Generative Models: Fingerprints

- Fingerprints are characterized by ridges and minutiae

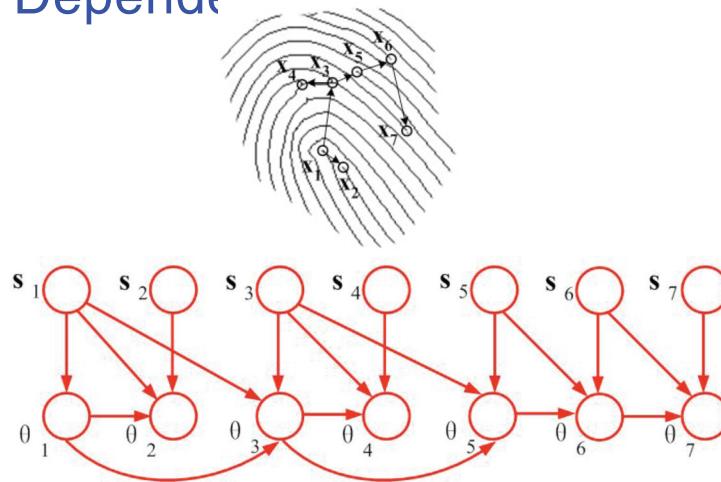


Generative Model: Minutiae

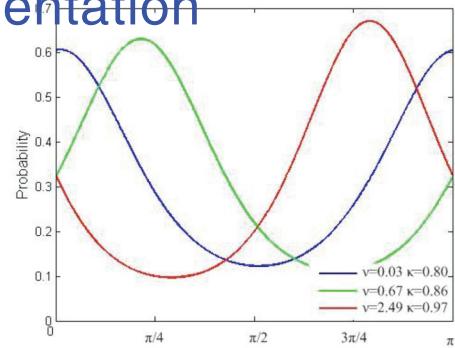
1. Distribution of Minutia Location



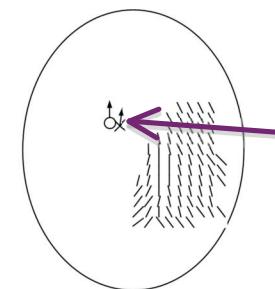
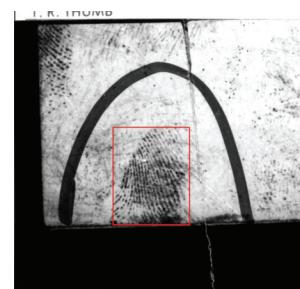
3. Distribution of Minutia Dependency



2. Distribution of Minutia Orientation

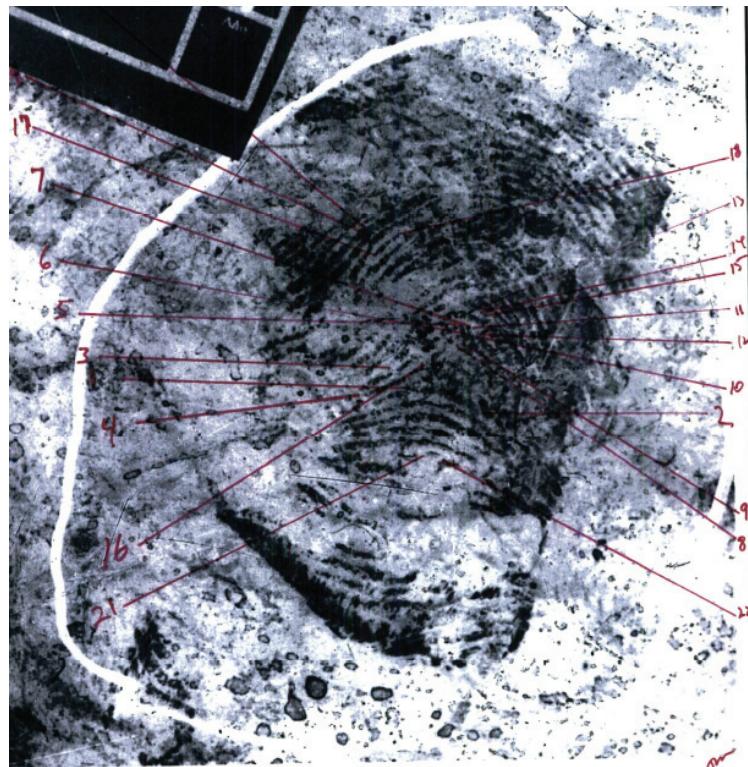


4. Distribution of Minutia Certainty



Core point
predicted
Using
Regression

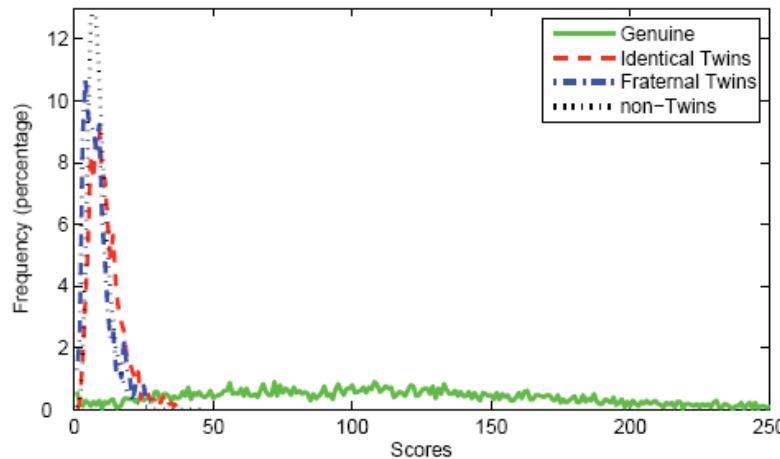
Madrid Bomber Case Revisited



- Madrid Bomber case 22 minutiae identified
- 10 were matched by three FBI experts
- Generative Model
 - 12 of 19 minutiae used for specific nPRC:
1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 17, and 18
 - specific nPRC
 - World population (6.8b) = 0.16
 - US population (330m) = 0.008

Discriminative Models: Fingerprints

- Need a similarity measure for fingerprints
- AFIS Minutiae Matcher
 - Score distributions give two probabilities
 - Likelihood Ratio
- Error Rates



	EER	EER threshold
Non-Twins	3.33%	19
Fraterna...	4.88%	23
Identical	5.09%	25

Discriminative Model: QD

Known

Nov. 10, 1999

From
Jim Elder
829 Loop Street, Apt. 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, kate Zack. It all started around six months ago while attending the "Ruby" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kates been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank You!

Jim

Letter shape

Questioned

November 10, 1999

From
Jim Elder
829 Loop Street, Apt. 300
Allentown, New York 14707

To Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, kate Zack.

It all started around six months ago while attending the "Ruby" Jazz concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

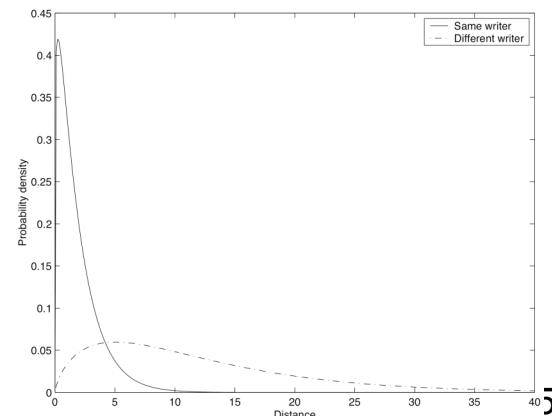
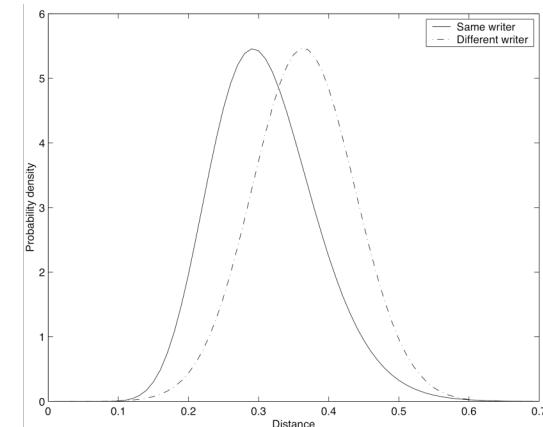
Bigram shape

Word shape

Handwriting Features and Similarities

Likelihood Ratio
obtained from
distributions of
similarities
for same and different
origination of evidence

Distributions of Similarities



Discriminative Model for Handwriting

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, we were told it was just exhaustion.

Kate has been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!

Jim

From Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

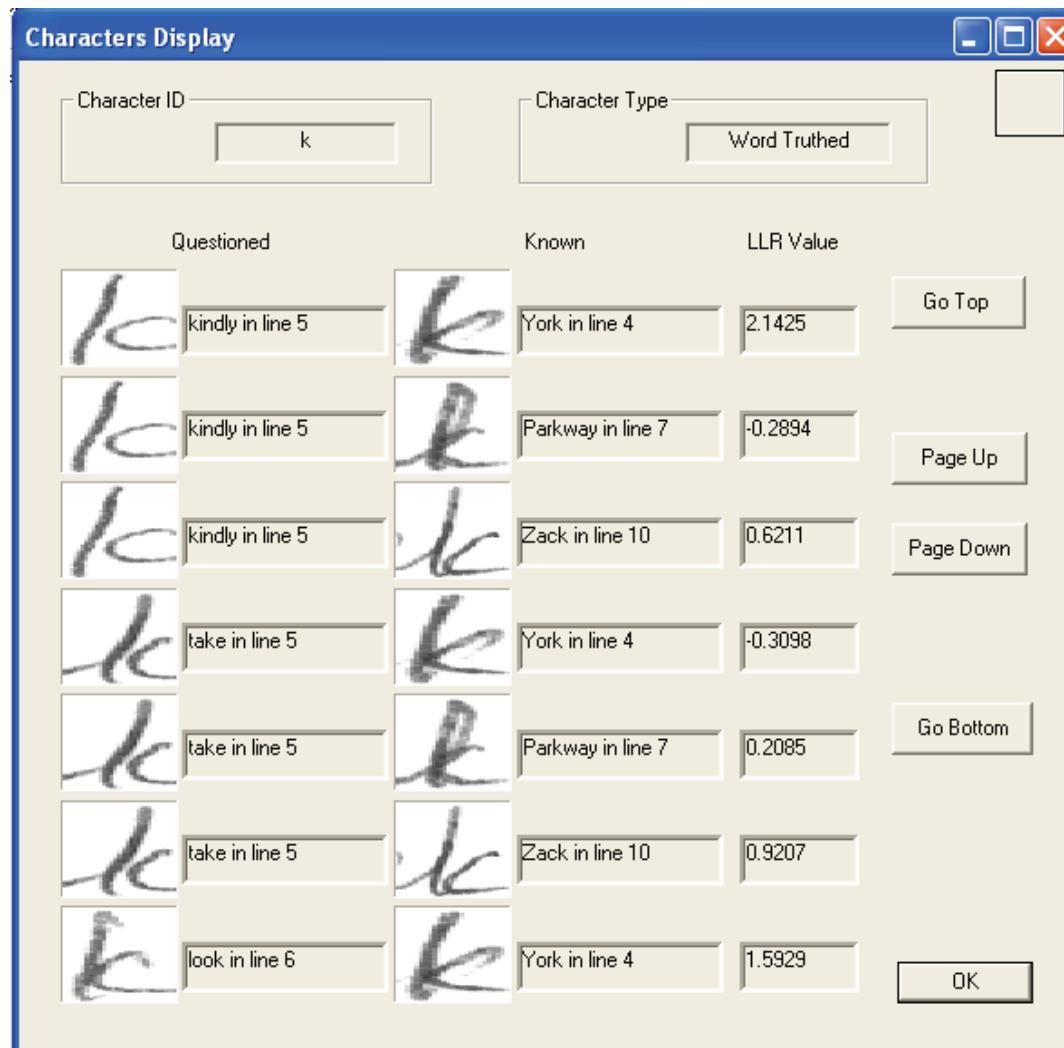
It all started around six months ago while attending the "Ruby" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked.

LLR Value for this pair is 41.52



1. Identified as same
2. Highly probable same
3. Probably same
4. Indicating same
5. No conclusion
6. Indicating different
7. Probably different
8. Highly probably different
9. Identified as different

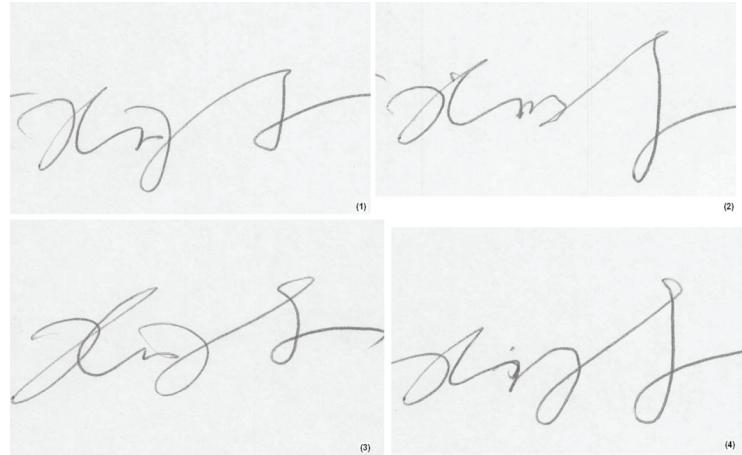
Human-Machine Interface: Dialog for character image comparison



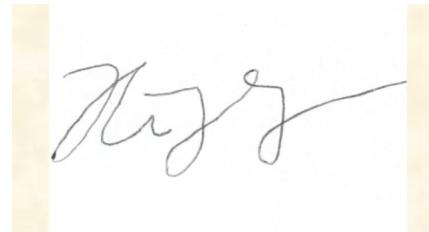
Signatures: Bayesian Adaptation

Wide variability and small learning sets for case at hand

Known Signatures



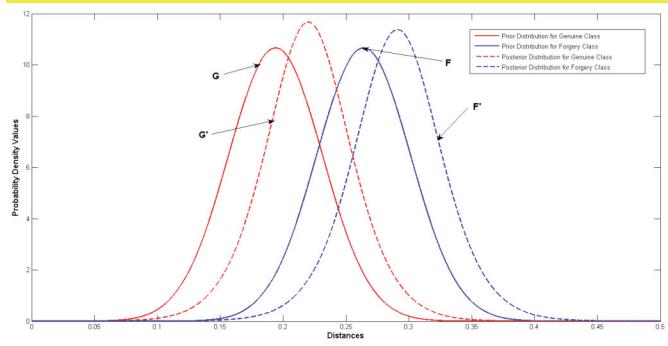
Questioned Signature



Approach:

Learn hyper-parameters from large data set
Adapt parameters using Bayesian approach

Bayesian Adaptation



	Error Rate
Only Prior	18%
Bayesian Adaptation	11.5% - 14.5%

	After Learning	After Adaptation
Parameter	$\theta_g \sim N(\mu_{g0}, \tau_{g0}^2)$ $\theta_f \sim N(\mu_{f0}, \tau_{f0}^2)$	$P(\theta_g Y_g) \propto N(\theta_g \mu_{g0}, \tau_g^2) \times N(Y_g \theta_g, \sigma_g^2) = N(\mu_g, \tau_g^2)$ $P(\theta_f Y_f) \propto N(\theta_f \mu_{f0}, \tau_f^2) \times N(Y_f \theta_f, \sigma_f^2) = N(\mu_f, \tau_f^2)$
Distance	$P(d_g G) = \int_{-\infty}^{\infty} P(d_g \theta_g) \times P(\theta_g Y_g) d\theta_g = \frac{1}{\sqrt{2\pi(\hat{\sigma}_g^2 + \tau_{g0}^2)}} e^{-\frac{(d_g - \mu_{g0})^2}{2(\hat{\sigma}_g^2 + \tau_{g0}^2)}}$ $P(d_f F) = \int_{-\infty}^{\infty} P(d_f \theta_f) \times P(\theta_f Y_f) d\theta_f = \frac{1}{\sqrt{2\pi(\hat{\sigma}_f^2 + \tau_{f0}^2)}} e^{-\frac{(d_f - \mu_{f0})^2}{2(\hat{\sigma}_f^2 + \tau_{f0}^2)}}$	$p(t G) = \int_{-\infty}^{\infty} p(t \theta_g) \times p(\theta_g Y_g) d\theta_g = \frac{1}{\sqrt{2\pi(\sigma_g^2 + \tau_g^2)}} e^{-\frac{(t - \mu_g)^2}{2(\sigma_g^2 + \tau_g^2)}}$ $p(t F) = \int_{-\infty}^{\infty} p(t \theta_f) \times p(\theta_f Y_f) d\theta_f = \frac{1}{\sqrt{2\pi(\sigma_f^2 + \tau_f^2)}} e^{-\frac{(t - \mu_f)^2}{2(\sigma_f^2 + \tau_f^2)}}$

Summary and Conclusion

- Uncertainty is easily expressed for DNA
 - Discrete counts
- For Impression Evidence it is much harder
 - Continuous features/tolerances
- New Models proposed (being validated)
 - Generative
 - PRC
 - Fingerprints, Handwriting
 - Discriminative (similarity automatically computed)
 - Likelihoods/ Bayesian scores